Two-Stage Subspace Constrained Precoding in Massive MIMO Cellular Systems

An Liu, *Member IEEE*, and Vincent Lau, *Fellow IEEE*,
Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology

***Abstract*—We propose a subspace constrained precoding scheme that exploits the spatial channel correlation structure in massive MIMO cellular systems to fully unleash the tremendous gain provided by massive antenna array with reduced channel state information (CSI) signaling overhead. The MIMO precoder at each base station (BS) is partitioned into an *inner precoder* and a *Transmit (Tx) subspace control* matrix. The inner precoder is adaptive to the local CSI at each BS for spatial multiplexing gain. The Tx subspace control is adaptive to the channel statistics for inter-cell interference mitigation and Quality of Service (QoS) optimization. Specifically, the Tx subspace control is formulated as a QoS optimization problem which involves an SINR chance constraint where the probability of each user's SINR not satisfying a service requirement must not exceed a given outage probability. Such chance constraint cannot be handled by the existing methods due to the two-stage precoding structure. To tackle this, we propose a *bi-convex approximation approach*, which consists of three key ingredients: random matrix theory, chance constrained optimization and semidefinite relaxation. Then we propose an efficient algorithm to find the optimal solution of the resulting bi-convex approximation problem. Simulations show that the proposed design has significant gain over various baselines.**

***Index Terms*—Massive MIMO, Subspace constrained precoding, QoS Guarantees**

I. Introduction

Massive MIMO has been regarded as one of the key technologies in future wireless networks. In massive MIMO cellular systems, each BS is equipped with $M \gg 1$ antennas. This large spatial degree of freedom (DoF) of massive MIMO systems can be exploited to significantly increase the spectrum and energy efficiency [1]. Specifically, there are two important roles for the spatial DoF introduced by the massive MIMO, namely the *intra-cell spatial multiplexing* and the *inter-cell interference mitigation*. However, there are several practical issues towards achieving the huge performance gain predicted by the theoretical analysis in massive MIMO systems. First, to realize the *spatial multiplexing gains* (combating intra-cell interference) within each BS using conventional MU-MIMO precoding [2], [3], real-time CSIT (i.e, channel state information at the BS) is required. In most of the existing works, Time-Division Duplex (TDD) is assumed and channel reciprocity can be exploited to obtain CSIT via uplink pilot training. However, in this paper, we are focusing on a more challenging but important Frequency-Division Duplex (FDD) mode of Massive MIMO. In this case, channel reciprocity cannot work and the channel estimation has to be obtained via downlink channel estimation and channel feedback, which is practically infeasible because the number of independent pilot symbols available for channel estimation is fundamentally limited by the channel coherence time and it may become much smaller than M as M grows large. Second, to realize the *inter-cell interference mitigation gains* in massive MIMO using coordinated beamforming [4] or cooperative MIMO [5], *real-time global CSIT* knowledge is required for cooperative MIMO and at least partial global CSIT (i.e., each BS needs to know the channels from this BS to all users) is required for coordinated beamforming. However, the acquisition of (partial) global CSIT is a very challenging problem in practical massive MIMO systems because both the downlink pilot training and the uplink CSI feedback overhead will become unacceptable as the number of antennas M grows large. It is even more difficult to obtain real-time (partial) global CSIT due to the backhaul signaling latency.

In this paper, we propose a two stage *subspace constrained precoding* for massive MIMO cellular systems. In practice, due to local scattering effects [6], [7], the channel vectors of users are usually concentrated in a subspace with a much smaller dimension than M when M is large. The proposed two-stage precoding takes advantage of this limited scattering property to achieve *intra-cell spatial multiplexing* and *inter-cell interference mitigation* without all the aforementioned practical issues. Specifically, the MIMO precoder at each BS is partitioned into an *inner precoder* and a semi-unitary *Tx subspace control matrix*. The inner precoder is adaptive to real-time local CSIT per BS for intra-cell spatial multiplexing gain. The Tx subspace control matrix is adaptive to long-term channel statistics to achieve the best tradeoff between direct link diversity gain and cross link (inter-cell) interference mitigation such that the QoS of the users is maximized. Such subspace constrained precoding structure simultaneously resolves the aforementioned practical challenges. For instance, the issue of insufficient pilot symbols for real-time local CSI estimation is resolved because the BS only needs to estimate the CSI within the subspace determined by the Tx subspace control, which is of a much smaller dimension than M. Furthermore, the Tx subspace control is adaptive to the long-term channel statistics, which is insensitive to the backhaul latency. As a result, the proposed subspace constrained precoding fully exploits the large number of antennas to simultaneously achieve intra-cell spatial multiplexing per BS and inter-cell interference mitigation without an expensive backhaul signaling requirement.

In [8], a similar two-stage precoding structure was proposed for single cell massive MIMO systems, where a simple pre-beamforming (with a pre-determined dimension) is used to control intra-cell interference and achieve spatial multiplexing gain based on block diagonalization (BD). However, the solution in [8] for single cell systems is fundamentally

	Proposed two stage precoding	**Two stage precoding in** [8]
System topology	Multi-cell massive MIMO system	Single cell massive MIMO system
Roles of two-stage precoding	Intra-cell & Inter-cell interference mitigation	Intra-cell interference mitigation
Subspace dimension partitioning	Dynamic	Static
Adaptive to heterogeneous path loss and QoS requirements explicitly	Yes	No
Design approach	Non-trivial design based on optimization	Simple design based on BD method
Performance	Good	Moderate

Table I: Summary of the differences compared with existing (state-of-the-art) two stage precoding scheme in [8].

different from the proposed solution for multi-cell systems with heterogeneous path loss and heterogeneous QoS requirements, as summarized in Table I. For example, the BD-based pre-beamforming design in [8] does not explicitly take into account the heterogeneous path loss and QoS requirements and thus may be far from optimal as shown in the simulations. Moreover, the pre-beamforming design in [8] is based on a heuristic (BD) method and the dimensions of the pre-beamforming matrices for different user clusters are pre-determined and fixed. In this paper, the design of Tx subspace control is formulated as a QoS optimization problem with individual QoS requirements and the dimensions of Tx subspace control matrices are also included in the optimization. As shown in the simulations, in multi-cell systems with heterogeneous path loss and QoS requirements, it is very important to dynamically allocate the dimensions of the Tx subspace control matrices over different user clusters. With optimal dynamic subspace dimension partitioning, the proposed solution can achieve a much better performance than the two-stage precoding in [8] with static subspace dimension partition. However, the optimization based solution in this paper also causes several non-trivial technical challenges.

- **SINR Chance Constraint under Two-Stage Precoding:** The QoS optimization problem involves an SINR chance constraint where the probability of each user's SINR not satisfying a service requirement must not exceed a given outage probability. Such SINR chance constraint does not have a closed form expression. There are various *Bernstein techniques* to derive a safe convex approximation for the chance constraints of standard forms [9], [10]. However, due to the two-stage precoding structure, the associated SINR chance constraint does not belong to any of the standard forms.
- **Combinatorial Nature of Dimension Partitioning:** Due to the combinatorial nature, the optimization of subspace dimension partitioning is very challenging and brute force solution has high complexity for large M.

In this paper, we address the above technical challenges using a novel *bi-convex approximation approach*. The first challenge is addressed using the *interference approximation* and *SINR chance constraint restriction*, where we combine the techniques in random matrix theory and chance constraint optimization to construct a deterministic and (asymptotically) conservative approximation for the SINR chance constraint. The second challenge is addressed using the *semidefinite relaxation (SDR)*, where we apply the SDR technique to obtain a relaxed bi-convex problem which does not explicitly involve the combinatorial optimization w.r.t. the dimension partitioning variable. We further prove the tightness of the SDR using the specific structure of the Tx subspace control problem. Finally, we propose a low complexity iterative algorithm to solve this bi-convex problem and find the Tx subspace control solution by combining the convex optimization method and the bisection search method. Simulations show that the proposed design achieves significant performance gain compared with various state-of-the-art baselines under various backhaul signaling latency.

Notations: The superscripts $(\cdot)^T$ and $(\cdot)^\dagger$ denote transpose and Hermitian respectively. For a set $\mathcal{S}$, $|\mathcal{S}|$ denotes the cardinality of $\mathcal{S}$. The operator $\text{diag}\,(\mathbf{a})$ represents a diagonal matrix whose diagonal elements are the elements of vector $\mathbf{a}$. For a set of K vectors $\mathbf{a}_i \in \mathbb{C}^M, i = 1, ..., K$ and a subset $\mathcal{S} \subseteq \{1, ..., K\}$, the notation $[\mathbf{a}]_{i\in\mathcal{S}}$ denote a $M \times |\mathcal{S}|$ matrix whose columns are drawn from the vectors in $\{\mathbf{a}_i, i \in \mathcal{S}\}$. The notation $\mathbb{U}^{M\times N}$ denotes the set of all $M \times N$ semi-unitary matrices. For a matrix $\mathbf{A}$, $\text{diag}\,(\mathbf{A})$ represents a diagonal matrix whose diagonal elements are the diagonal elements of $\mathbf{A}$. $\text{span}\,(\mathbf{A})$ represents the subspace spanned by the columns of a matrix $\mathbf{A}$. $\|\mathbf{A}\|$ is the spectral radius of $\mathbf{A}$.

II. System Model

A. Massive MIMO Cellular System

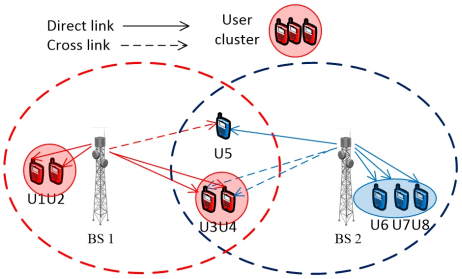
Consider the downlink of a massive MIMO cellular system with L BSs and K single-antenna users as illustrated in Fig. 1 for $L = 2$ and $K = 8$. Each BS has M antennas with M much larger than the number of the associated users. The massive MIMO cellular system can be represented by a topology graph as define below.

Definition 1 (System Topology Graph)**.** Define the *topology graph* of the massive MIMO cellular system as a bipartite graph $\mathcal{G}_T = \{\mathcal{B}, \mathcal{U}, \mathcal{E}\}$, where $\mathcal{B}$ denotes the set of all BS nodes, $\mathcal{U}$ denotes the set of all user nodes, and $\mathcal{E}$ is the set of all edges between the BSs and users. An edge $(k, l) \in \mathcal{E}$ between BS node $l \in \mathcal{B}$ and user node $k \in \mathcal{U}$ represents a wireless link between them. Each edge $(k, l) \in \mathcal{E}$ is associated with a CSI label $\mathbf{h}_{k,l} \in \mathbb{C}^M$, which represents the channel vector between BS l and user k. For each user node k, let b_k denote the serving BS. ■

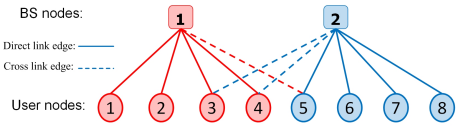
An example of a massive MIMO cellular system and the corresponding topology graph is illustrated in Fig. 1.

B. Massive MIMO Channel Model

The channel from BS l to user k is modeled as $\mathbf{h}_{k,l} = \mathbf{\Theta}_{k,l}^{1/2}\mathbf{h}_{k,l}^w,\ \forall k, l$, where $\mathbf{h}_{k,l}^w \in \mathbb{C}^M$ has i.i.d. complex entries



(a) A massive MIMO cellular system with 2 BSs and 8 users.



(b) The corresponding topology graph $\mathcal{G}_T(\mathbf{\Theta}) = \{\mathcal{B}, \mathcal{U}, \mathcal{E}\}$, where $\mathcal{B} = \{1, 2\}$, $\mathcal{U} = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

Figure 1: An example of a massive MIMO cellular system and the corresponding topology graph.

of zero mean and unit variance; and $\mathbf{\Theta}_{k,l} \in \mathbb{C}^{M\times M}$ is a positive semi-definite spatial correlation matrix between BS l and user k. Assume block fading channel where $\mathbf{h}_{k,l}^w$ is fixed for a time slot but changes over time slots. As such, the CSI is divided into instantaneous CSI $\mathbf{H} = \{\mathbf{h}_{k,l}\}$ and global statistical information $\mathbf{\Theta} \triangleq \{\mathbf{\Theta}_{k,l}\}$ (spatial correlation matrices). We make the following assumption on the spatial correlation matrices.

Assumption 1 (Spatial Correlation Model). The spatial channel correlation matrices satisfy the following two assumptions.

1) The number of dominant eigenvalues (i.e., the eigenvalues that are comparable to the maximum eigenvalue) of $\mathbf{\Theta}_{k,l,}, \forall k, l$ is relatively small compared to the number of antennas M.
2) The users can be grouped into N user clusters $\mathcal{U}_n, n = 1, ..., N$ such that all users in $\mathcal{U}_n$ are associated with the same BS denoted by $\bar{l}_n$ and $\mathbf{\Theta}_{k,\bar{l}_n} = L_{k,\bar{l}_n}\mathbf{\Theta}_n^\circ, \forall k \in \mathcal{U}_n$, where $L_{k,\bar{l}_n} > 0$ is the path loss between BS $\bar{l}_n$ and user k, and $\mathrm{Tr}(\mathbf{\Theta}_n^\circ) = M$ is a normalized spatial correlation matrix.

Many experimental measurements as reported in [11], [12] show that practical MIMO channels indeed have spatial channel correlation due to limited scattering or LOS transmission. The standard channel models such as IMT-advanced channel model [13] also predicts that MIMO channel can have high spatial correlation especially when the angular spread is small. Furthermore, in a massive MIMO system with a large number of antennas assembled within a limited space at the BS, the channels are more likely to be highly correlated due to insufficient spacing among the antennas [1], [14]. As a result, the first assumption has been made in many existing works on massive MIMO systems, see, e.g., [8], [14]–[18]. In practice, we can also cluster users into groups using the user clustering algorithm in [18] such that the second assumption can be closely achieved [8]. Note that the above channel model is more general than that considered in [8]. For example, the n-th cluster $\mathcal{U}_n$ is allowed to contain a single user and the users in the same cluster can have different path loss.

At each time slot, linear precoding is employed at BS $\bar{l}_n$ to support simultaneous downlink transmissions to the associated users in $\mathcal{U}_n$. Let $\overline{n}_k$ denote the index of the cluster that contains user k and let $\mathcal{B}_k = \left\{n : n \neq \overline{n}_k, \left(k, \bar{l}_n\right) \in \mathcal{E}\right\}$ denote the set of user clusters that interfere with user k. Then the received signal for user k can be expressed as:

$$y_k = \mathbf{h}_{k,b_k}^\dagger \sqrt{P_k}\mathbf{v}_k s_k + \underbrace{\mathbf{h}_{k,b_k}^\dagger \sum_{k'\in\mathcal{U}_{\overline{n}_k}\setminus\{k\}} \sqrt{P_{k'}}\mathbf{v}_{k'} s_{k'}}_{\text{intra-cluster interference}} + \underbrace{\sum_{n\in\mathcal{B}_k} \mathbf{h}_{k,\bar{l}_n}^\dagger \mathbf{V}_n \mathbf{P}_n^{1/2}\mathbf{s}_n}_{\text{inter-cluster interference}} + z_k, \quad (1)$$

where $s_k \sim \mathcal{CN}(0,1)$ is the data symbol for user k; $\mathbf{v}_k \in \mathbb{C}^M$ with $\|\mathbf{v}_k\| = 1$ is the precoding vector for user k; P_k is the power allocated to user k; $\mathbf{s}_n = [s_l]_{l\in\mathcal{U}_n} \in \mathbb{C}^{|\mathcal{U}_n|}$ is the data symbol vector at BS n; $\mathbf{V}_n = [\mathbf{v}_l]_{l\in\mathcal{U}_n} \in \mathbb{C}^{M\times|\mathcal{U}_n|}$ is the precoding matrix at BS n; $\mathbf{P}_n = \mathrm{diag}\left([P_l]_{l\in\mathcal{U}_n}\right) \in \mathbb{R}_+^{|\mathcal{U}_n|\times|\mathcal{U}_n|}$ is the power allocation matrix for the n-th user cluster; and $z_k \sim \mathcal{CN}(0,1)$ is the AWGN noise.

C. Two-stage Subspace Constrained Precoding

The precoder $\mathbf{V}_n$ for the n-th cluster at BS $\bar{l}_n$ is assumed to have a two-stage structure $\mathbf{V}_n = \mathbf{F}_n\mathbf{G}_n$. The *Tx subspace control variable* $\mathbf{F}_n \in \mathbb{U}^{M\times S_n}$ is used to mitigate the inter-cluster interference (including inter-cell interference) in massive MIMO systems, where $S_n \in \{|\mathcal{U}_n|, ..., M\}$ is called the *dimension partitioning variable* for the n-th cluster. For convenience, define $\mathbf{F} = \{\mathbf{F}_1, ..., \mathbf{F}_N\}$ as the set of Tx subspace control variables for all user clusters. Both $\mathbf{F}$ and $S_n, \forall n$ are computed at a *central node* based on the global statistical information $\mathbf{\Theta}$. The inner precoder $\mathbf{G}_n = [\mathbf{g}_k]_{k\in\mathcal{U}_n} \in \mathbb{C}^{S_n\times|\mathcal{U}_n|}$ is used to realize the intra-cluster spatial multiplexing gain using simple zero-forcing (ZF) precoding. Specifically, for a user $k \in \mathcal{U}_n$, define $\widetilde{\mathbf{h}}_{k,n} \triangleq \mathbf{F}_n^\dagger \mathbf{h}_{k,\bar{l}_n} \in \mathbb{C}^{S_n}$ as its *effective channel*. Then, the ZF inner precoding vector $\mathbf{g}_k$ with $\|\mathbf{g}_k\| = 1$ is obtained by the projection of the effective channel $\widetilde{\mathbf{h}}_{k,n}$ on the orthogonal complement of the subspace span $\left(\left[\widetilde{\mathbf{h}}_{k',n}\right]_{k'\in\mathcal{U}_n\setminus\{k\}}\right)$. Hence, $\mathbf{G}_n$ is computed locally at BS $\bar{l}_n$ based on the local real-time instantaneous CSI $\widetilde{\mathbf{H}}_n = \left[\widetilde{\mathbf{h}}_{k,n}\right]_{k\in\mathcal{U}_n}^\dagger \in \mathbb{C}^{|\mathcal{U}_n|\times S_n}$. Note that in the proposed two stage precoding, only the inner precoder is based on ZF techniques for analytical tractability. The Tx subspace control is not based on ZF techniques but based on the QoS optimization problem $\mathcal{P}$ in Section III.

The diversity gain of the users in the n-th cluster increases with the dimension partitioning variable S_n. On the other hand, the interference leakage to other user clusters also increases with S_n. As a result, there is a fundamental tradeoff between direct link diversity gain and cross link interference leakage. It is important to dynamically adapt the dimension

partitioning variable S_n based on the spatial channel correlation matrices $\mathbf{\Theta}$ to optimize the overall performance. In general, the simple BD-based subspace precoder $\mathbf{F}_n$ may be far from optimal as will shown in the simulations. We shall formulate the design of the Tx subspace control $\mathbf{F}$ and *subspace dimension partitioning* $S_n, \forall n$ formally in Section III.

III. Optimization Formulation for Tx Subspace and Dimension Partitioning

Assume that user k has perfect knowledge of the effective single-input single-output (SISO) channel $\mathbf{h}_{k,b_k}^\dagger \mathbf{F}_{\overline{n}_k} \mathbf{g}_k$ and the interference-plus-noise power. Then for given Tx subspace control $\mathbf{F}$, subspace dimensions $\{S_n\}$, and instantaneous CSI $\mathbf{H}$, the SINR of user k is

$$\text{SINR}_k = \frac{P_k \left| \mathbf{h}_{k,b_k}^\dagger \mathbf{F}_{\overline{n}_k} \mathbf{g}_k \right|^2}{\sum_{n \in \mathcal{B}_k} \mathbf{h}_{k,\bar{l}_n}^\dagger \mathbf{F}_n \mathbf{G}_n \mathbf{P}_n \mathbf{G}_n^\dagger \mathbf{F}_n^\dagger \mathbf{h}_{k,\bar{l}_n} + 1}. \quad (2)$$

where $\mathbf{G}_n$ is the ZF inner precoder given by a function of $\mathbf{F}_n$ and $\mathbf{H}$ (or more precisely, the effective channels $\widetilde{\mathbf{H}}_n$). In this paper, we focus on designing Tx subspace control $\mathbf{F}$ to optimize the QoS of the users. Specifically, we consider the following QoS optimization problem:

$$\mathcal{P}: \max_{\mathbf{F}, \{S_n\}, \gamma \geq \gamma^\circ} \gamma$$

$$\text{s.t.} \quad \Pr\{\text{SINR}_k \geq w_k \gamma\} \geq 1 - \epsilon_k, \forall k; \quad (3)$$

$$\mathbf{F}_n \in \mathbb{U}^{M \times S_n}, S_n \in \{|\mathcal{U}_n|, ..., M\}, \forall n, \quad (4)$$

where the SINR chance constraint in (3) ensures that the probability of SINR_k not satisfying a service requirement $w_k\gamma$ is less than a maximum allowable outage probability $\epsilon_k \in (0,1)$, and $w_k > 0$ is the QoS weight for user k. The constraint $\gamma \geq \gamma^\circ$ is used to guarantee some minimum QoS for all users. The constraint $S_n \geq |\mathcal{U}_n|$ is to ensure that the ZF inner precoding is feasible at each BS. The QoS constraint in (3) is especially useful for media streaming applications where there is a stringent delay requirement for the media packets requested by the users and thus it is desirable to maintain a fixed data rate (with high probability) for each user. The QoS weights w_k's and the maximum allowable outage probabilities ϵ_k's can be used to provide differential QoS for different users.

Note that the optimization of the Tx subspace control variable $\mathbf{F}_n$ in Problem $\mathcal{P}$ includes the optimization of both the dimension partitioning variable S_n (or equivalently, the dimension of $\mathbf{F}_n$) and the value of $\mathbf{F}_n$. Both $\mathbf{F}_n$'s and S_n's are adaptive to the spatial channel correlation matrices $\mathbf{\Theta}$, i.e., they are fixed and independent of the instantaneous CSI $\mathbf{H}$ for given $\mathbf{\Theta}$.

Remark 2 (Admission Control). Note that Problem $\mathcal{P}$ may not always be feasible, i.e., the minimum QoS requirement $w_k\gamma^\circ$ cannot be guaranteed for all users even with infinite transmit power. In practice, an admission control can be used to ensure that there is enough resource in the system to guarantee the minimum QoS for all users. The admission control is a challenging problem and a systematic design of admission control is out of the scope of this paper. However, we can use a simple admission control based on the solution of Problem $\mathcal{P}$. For example, whenever a new user arrives, the admission control solves Problem $\mathcal{P}$ with all users (including the newly arrival user). If Problem $\mathcal{P}$ is feasible, this user is admissible. Otherwise, this user is not allowed to access the system. As a result, we can focus on the case when Problem $\mathcal{P}$ is feasible.

Problem $\mathcal{P}$ is a very challenging problem. First, the probability functions in the SINR chance constraint (3) do not have closed form expressions. Thus, one may only resort to approximation methods. Due to the two-stage precoding structure, the inner ZF precoder $\mathbf{G}_n$ is also a function of the Tx subspace control variable $\mathbf{F}_n$. As a result, the existing approximation methods for the "standard form" chance constraints in [9], [10] cannot be applied here. Second, the dimension partitioning variable S_n also needs to be optimized, which is a difficult combinatorial problem. Third, the domain $\mathbb{U}^{M \times S_n}$ of $\mathbf{F}_n$ is non-convex. To tackle the above challenges, we propose a novel *bi-convex approximation approach*, which consists of three key ingredients: random matrix theory, chance constrained optimization and semidefinite relaxation (SDR). Then we propose an efficient algorithm to find the optimal solution of the resulting bi-convex approximation problem.

IV. The Bi-convex Approximation Approach

Since it is difficult to solve Problem $\mathcal{P}$ directly, we propose to find a sub-optimal solution by solving a bi-convex approximation of $\mathcal{P}$. The proposed bi-convex approximation contains three steps, where each step solves one key technical challenge associated with Problem $\mathcal{P}$.

A. Interference Approximation Step

To handle the SINR chance constraint in (3), we need to find a simple characterization of the distribution of SINR_k, which is very difficult because the interference term $I_k \triangleq \sum_{n \in \mathcal{B}_k} \mathbf{h}_{k,\bar{l}_n}^\dagger \mathbf{F}_n \mathbf{G}_n \mathbf{P}_n \mathbf{G}_n^\dagger \mathbf{F}_n^\dagger \mathbf{h}_{k,\bar{l}_n}$ in SINR_k depends on the instantaneous channel vectors $\mathbf{h}_{k,\bar{l}_n}, n \in \mathcal{B}_k$ of all the cross links of user k. To simplify the characterization of SINR_k, the interference approximation step is used to address the following challenge.

Challenge 1 (Deterministic Approximation of I_k). Find a deterministic approximation $\hat{I}_k$ of I_k such that $\hat{I}_k$ does not depend on the instantaneous channel vectors $\mathbf{h}_{k,\bar{l}_n}, n \in \mathcal{B}_k$ of the cross links of user k.

We resort to the random matrix theory to solve the above challenge. Specifically, we apply the random matrix theory to derive an asymptotic (and deterministic) upper bound $\hat{I}_k$ of I_k as $M \to \infty$ and $|\mathcal{U}_n| \to \infty, \forall n$. We then use $\hat{I}_k$ as an approximation of I_k for finite but large M and $|\mathcal{U}_n|$'s. Throughout the paper, the notation $M \to \infty$ refers to $M \to \infty$ and $|\mathcal{U}_n| \to \infty, \forall n$ such that $0 < \liminf_{M\to\infty} |\mathcal{U}_n|/M \leq \limsup_{M\to\infty} |\mathcal{U}_n|/M < \infty$. The following assumptions are required in order to derive the asymptotic upper bound of I_k.

Assumption 2 (Technical Assumptions for Interference Approximation). All spatial correlation matrices $\mathbf{\Theta}_{k,l}, \forall k, l$ have

uniformly bounded spectral norm on M, i.e.,

$$\limsup_{M\to\infty}\sup_{1\le k\le K}\|\mathbf{\Theta}_{k,l}\| \quad < \quad \infty,\ \forall l. \tag{5}$$

Assumption 2 is satisfied by many MIMO channel model and it is a standard assumption in the literature, see e.g., [15], [16].

Under Assumption 2, we have the following theorem.

Theorem 3 (Asymptotic Interference Upper Bound)**.** *Under Assumption 2 and a feasible Tx subspace control* $\mathbf{F}$ *(i.e.,* $\mathbf{F}$ *satisfies the constraints in (3) and (4) for some* $\gamma \ge \gamma^{\circ}$*), we have* $I_k - \hat{I}_k \overset{a.s}{\le} 0$ *as* $M \to \infty$*, where*

$$\hat{I}_k \triangleq \sum_{n\in\mathcal{B}_k} \overline{P}_n Tr\left(\mathbf{F}_n^{\dagger}\mathbf{\Theta}_{k,\bar{l}_n}\mathbf{F}_n\right).$$

and $\overline{P}_n = \sum_{k'\in\mathcal{U}_n} P_{k'}$ *is the sum transmit power of the users in the* n*-th cluster.*

Please refer to Appendix A for the proof.

Replacing the interference term I_k in SINR_k using the asymptotic interference upper bound $\hat{I}_k$ in Theorem 3, we obtain an asymptotically safe approximation of the SINR chance constraint in (3)[1]:

$$\Pr\left\{\left|\mathbf{h}_{k,b_k}^{\dagger}\mathbf{F}_{\overline{n}_k}\mathbf{g}_k\right|^2 \ge \frac{w_k\gamma}{P_k}\left(\hat{I}_k+1\right)\right\} \ge 1-\epsilon_k. \tag{6}$$

B. SINR Chance Constraint Restriction Step

The asymptotically safe approximate SINR chance constraint in (6) remains intractable although it appears to be relatively easier to handle than the original counterparts in (3). The restriction step aims to find a quadratic conservative approximation of (6). There are some existing methods that use the worst-case deterministic constraints to approximate various forms of chance constraints [9], [10], [19]. For example, in the context of transmit beamforming design for MU-MIMO downlink with imperfect CSI [10], the Bernstein-type inequality was used to construct a convex restriction of the SINR chance constraint that involves a quadratic function of the standard complex Gaussian vector. Unlike the conventional transmit beamforming problem, in our problem, the MIMO precoder is divided into Tx subspace control $\mathbf{F}_n$ and inner precoder $\mathbf{G}_n$ which change at different time scales. As a result, the asymptotically safe approximate SINR chance constraint in (6) does not belong to any of the standard forms that have been studied in the existing works. For example, the inner precoding vector $\mathbf{g}_k$ is also a function of the random channel vectors and thus $\left|\mathbf{h}_{k,b_k}^{\dagger}\mathbf{F}_{\overline{n}_k}\mathbf{g}_k\right|^2$ is not a quadratic function of the standard complex Gaussian vector. In our solution, the restriction step entails finding a solution to the following:

Challenge 2 (Asymptotically Quadratic Restriction of SINR Chance Constraint (3)). Find a quadratic function $\mathbf{f}_k(\mathbf{F})$ from $\mathbf{F}$ to $\mathbb{R}^2$ such that if $\mathbf{f}_k(\mathbf{F}) \ge \mathbf{0}$, then the SINR chance constraint in (3) is asymptotically satisfied as $M \to \infty$

[1]Asymptotically safe approximation means that if (6) is satisfied, then the SINR chance constraint in (3) is satisfied as $M \to \infty$.

Our strategy is as follows. First, we show that conditioned on $\widetilde{\mathbf{H}}_{-k} \triangleq \left[\widetilde{\mathbf{h}}_{k',\overline{n}_k}\right]_{k'\in\mathcal{U}_{\overline{n}_k}\setminus\{k\}}$, the effective channel gain $\left|\mathbf{h}_{k,b_k}^{\dagger}\mathbf{F}_{\overline{n}_k}\mathbf{g}_k\right|^2$ of user k can be expressed as a quadratic function of a standard complex Gaussian vector. Then we apply the standard method in [9], [10] to construct a quadratic restriction of the chance constraint in (6) conditioned on a given $\widetilde{\mathbf{H}}_{-k}$. Finally, we use the quadratic restriction of the conditional chance constraint of (6) with the "worst case" $\widetilde{\mathbf{H}}_{-k}$ as a quadratic restriction of the unconditional chance constraint in (6). Since (6) is an asymptotically safe approximation of (3), the result will then be an asymptotically quadratic restriction of the SINR chance constraint in (3). The final result is summarized in the following theorem. Please refer to Appendix B for the detailed proof.

Theorem 4 (Solution to Challenge 2)**.** *The asymptotically safe approximate SINR chance constraint in (6) is satisfied if*

$$\begin{aligned}&\left(1-\tfrac{\sqrt{2\delta_k}}{\sigma_k}\right)Tr\left(\overline{\mathbf{\Theta}}_{\overline{n}_k}^{\circ}\mathbf{F}_{\overline{n}_k}\mathbf{F}_{\overline{n}_k}^{\dagger}\right)\\ &\ge \gamma\textstyle\sum_{n\in\mathcal{B}_k} Tr\left(\overline{\mathbf{\Theta}}_{k,n}\mathbf{F}_n\mathbf{F}_n^{\dagger}\right)+c_k(\gamma),\\ &Tr\left(\overline{\mathbf{\Theta}}_{\overline{n}_k}^{\circ}\mathbf{F}_{\overline{n}_k}\mathbf{F}_{\overline{n}_k}^{\dagger}\right) \ge \sigma_k^2+|\mathcal{U}_{\overline{n}_k}|-1,\end{aligned} \tag{7}$$

where $\delta_k = \ln\frac{1}{\epsilon_k}$, $\sigma_k = \frac{\sqrt{2\delta_k}+\sqrt{2\delta_k+4}}{2}$, $\overline{\mathbf{\Theta}}_{\overline{n}_k}^{\circ} = \frac{1}{\lambda_{\overline{n}_k}}\mathbf{\Theta}_{\overline{n}_k}^{\circ}$, $\lambda_{\overline{n}_k}$ *is the largest eigenvalue of* $\mathbf{\Theta}_{\overline{n}_k}^{\circ}$, $\overline{\mathbf{\Theta}}_{k,n} = \frac{w_k\overline{P}_n}{\lambda_{\overline{n}_k}P_k}\mathbf{\Theta}_{k,\bar{l}_n}$ *and* $c_k(\gamma) = \frac{w_k\gamma}{\lambda_{\overline{n}_k}P_k} + \left(1-\frac{\sqrt{2\delta_k}}{\sigma_k}\right)(|\mathcal{U}_{\overline{n}_k}|-1)$.

In (7), $\text{Tr}\left(\overline{\mathbf{\Theta}}_{\overline{n}_k}^{\circ}\mathbf{F}_{\overline{n}_k}\mathbf{F}_{\overline{n}_k}^{\dagger}\right)$ and $\sum_{n\in\mathcal{B}_k}\text{Tr}\left(\overline{\mathbf{\Theta}}_{k,n}\mathbf{F}_n\mathbf{F}_n^{\dagger}\right)$ can be interpreted as the signal power of the direct link and interference leakage from the cross links. Hence, the quadratic restriction of (6) in (7) captures the key tradeoff between direct link signal power and cross link interference leakage.

Remark 5. The asymptotically quadratic restriction in (7) is also valid under finite M. The asymptotic approach based on random matrix theory has been widely used in wireless communications especially when it is difficult to directly analyze the original system and the validity of such approach has been verified by numerous works. In this paper, we utilize random matrix theory and obtain (7) as a safe approximation of the original SINR chance constraint when $M \to \infty$. It has been shown in [20] that the random matrix theory results are pretty good approximation even when M is small. Hence, (7) is a safe approximation of the original SINR chance constraint even under finite M in massive MIMO (as illustrated in Fig. 3 in Section VI).

According to Theorem 3 and Theorem 4, (7) is a quadratic restriction of the original SINR chance constraint in (3) for large M. Replacing the SINR chance constraint (3) in $\mathcal{P}$ with the quadratic constraint in (7), we have the following asymptotically conservative formulation of Problem $\mathcal{P}$:

$$\begin{aligned}\mathcal{P}_1:\ &\max_{\mathbf{F},\{S_n\},\gamma\ge\gamma^{\circ}}\ \gamma\\ \text{s.t.}\ \ &(7)\text{ is satisfied for all } k,\\ &\mathbf{F}_n\in\mathbb{U}^{M\times S_n}, S_n\in\{|\mathcal{U}_n|,...,M\},\forall n.\end{aligned} \tag{8}$$

C. Semidefinite Relaxation Step

By combining the techniques in random matrix theory and chance constraint optimization, we have constructed an asymptotically quadratic restriction for the complicated SINR chance constraint in (3). However, we still need to solve the challenge associated with the combinatorial optimization of the dimension partitioning variable S_n and the semi-unitary constraint on $\mathbf{F}_n$.

Challenge 3 (Dimension Partitioning and Semi-unitary Constraint). Find a bi-convex problem which is equivalent to Problem $\mathcal{P}_1$ and does not involve the combinatorial optimization w.r.t. the dimension partitioning variable S_n.

In the following, we apply SDR to solve the above challenge. It is easy to see that Problem $\mathcal{P}_1$ is equivalent to the following problem

$$\begin{aligned}\mathcal{P}_2: \ &\max_{\mathbf{W},\gamma\geq\gamma^\circ}\gamma\\ \text{s.t.}\quad &\left(1-\tfrac{\sqrt{2\delta_k}}{\sigma_k}\right)\mathrm{Tr}\left(\overline{\mathbf{\Theta}}_{\overline{n}_k}^{\circ}\mathbf{W}_{\overline{n}_k}\right)\\ &\geq\gamma\textstyle\sum_{n\in\mathcal{B}_k}\mathrm{Tr}\left(\overline{\mathbf{\Theta}}_{k,n}\mathbf{W}_n\right)+c_k\left(\gamma\right),\forall k; \qquad (9)\\ &\mathbf{W}_n=\mathbf{W}_n^2,\mathrm{Tr}\left(\overline{\mathbf{\Theta}}_n^{\circ}\mathbf{W}_n\right)\geq\eta_n,\mathbf{W}_n\succeq 0,\forall n, \qquad (10)\end{aligned}$$

where $\mathbf{W}=\{\mathbf{W}_1,...,\mathbf{W}_N\}$ with $\mathbf{W}_n=\mathbf{F}_n\mathbf{F}_n^\dagger\in\mathbb{C}^{M\times M},n=1,...,N$, and $\eta_n=\max_{k\in\mathcal{U}_n}\left(\sigma_k^2+|\mathcal{U}_n|-1\right)$. The constraint in (10) is derived from the constraint (8) in $\mathcal{P}_1$ and the second constraint in (7) using the fact that $\mathbf{W}_n$ can be expressed as $\mathbf{W}_n=\mathbf{F}_n\mathbf{F}_n^\dagger$ with $\mathbf{F}_n\in\mathbb{U}^{M\times S_n}$ if and only if $\mathbf{W}_n=\mathbf{W}_n^2,\mathrm{Tr}\left(\mathbf{W}_n\right)=S_n,\mathbf{W}_n\succeq 0$ (Note that $\mathrm{Tr}\left(\overline{\mathbf{\Theta}}_n^{\circ}\mathbf{W}_n\right)\geq\eta_n$ implies that $\mathrm{Tr}\left(\mathbf{W}_n\right)\geq|\mathcal{U}_n|$). For any feasible solution $\mathbf{W}$ of $\mathcal{P}_2$, we can calculate the corresponding feasible solution $\mathbf{F}=\{\mathbf{F}_1,...,\mathbf{F}_N\}$ and $S_n,\forall n$ of $\mathcal{P}_1$ using eigenvalue decomposition (EVD) $\mathbf{W}_n=\mathbf{F}_n\mathbf{I}_{S_n}\mathbf{F}_n^\dagger,\forall n$ and $S_n=\mathrm{Tr}\left(\mathbf{W}_n\right),\forall n$ respectively.

Problem $\mathcal{P}_2$ is still difficult to solve due to the non-convex constraint $\mathbf{W}_n=\mathbf{W}_n^2$ in (10) and the bi-linear term $\gamma\sum_{n\in\mathcal{B}_k}\mathrm{Tr}\left(\overline{\mathbf{\Theta}}_{k,n}\mathbf{W}_n\right)$ in (9). To make the problem tractable, we apply SDR to obtain a convex relaxation of the constraint $\mathbf{W}_n=\mathbf{W}_n^2$ as $\mathbf{W}_n-\mathbf{W}_n^2\succeq\mathbf{0}$, which is further equivalent to the following convex constraint

$$\left[\begin{array}{cc}\mathbf{I}_M & \mathbf{W}_n\\ \mathbf{W}_n & \mathbf{W}_n\end{array}\right]\succeq\mathbf{0}. \qquad (11)$$

Then by replacing the constraint $\mathbf{W}_n=\mathbf{W}_n^2$ with the relaxed constraint in (11) and removing the constraint $\gamma\geq\gamma^\circ$, we obtain a relaxed problem of $\mathcal{P}_2$ as

$$\begin{aligned}\hat{\mathcal{P}}: \ &\max_{\mathbf{W},\gamma}\gamma\\ \text{s.t.}\quad &a_k\mathrm{Tr}\left(\overline{\mathbf{\Theta}}_{\overline{n}_k}^{\circ}\mathbf{W}_{\overline{n}_k}\right)\geq\\ &\gamma\textstyle\sum_{n\in\mathcal{B}_k}\mathrm{Tr}\left(\overline{\mathbf{\Theta}}_{k,n}\mathbf{W}_n\right)+c_k\left(\gamma\right),\forall k \qquad (12)\\ &\mathrm{Tr}\left(\overline{\mathbf{\Theta}}_n^{\circ}\mathbf{W}_n\right)\geq\eta_n,\forall n \qquad (13)\\ &\left[\begin{array}{cc}\mathbf{I}_M & \mathbf{W}_n\\ \mathbf{W}_n & \mathbf{W}_n\end{array}\right]\succeq\mathbf{0},\mathbf{W}_n\succeq 0,\forall n \qquad (14)\end{aligned}$$

where $a_k=\left(1-\frac{\sqrt{2\delta_k}}{\sigma_k}\right)$.

Problem $\hat{\mathcal{P}}$ is a bi-convex problem, i.e., it is convex w.r.t. $\mathbf{W}$ (γ) for fixed γ ($\mathbf{W}$), but is not jointly convex. Note that Problem $\hat{\mathcal{P}}$ does not involve the combinatorial optimization w.r.t. the dimension partitioning variable S_n because the optimal S_n is automatically determined by the rank of the optimal $\mathbf{W}_n$. To completely solve Challenge 3, we further prove the equivalence between $\hat{\mathcal{P}}$ and $\mathcal{P}_1$ in the following theorem.

Theorem 6 (Equivalence between $\hat{\mathcal{P}}$ and $\mathcal{P}_1$)**.** *The optimal solution $\mathbf{W}^\star,\gamma^\star$ of $\hat{\mathcal{P}}$ satisfies $\mathbf{W}_n^\star-\mathbf{W}_n^{\star 2}=\mathbf{0},\forall n$. Moreover, if $\mathcal{P}_1$ is feasible, then $\gamma^\star\geq\gamma^\circ$ and thus $\mathbf{W}^\star,\gamma^\star$ is also optimal for $\mathcal{P}_1$.*

Please refer to Appendix C for the proof.

D. Optimal Solution of Problem $\hat{\mathcal{P}}$

In this section, we propose an iterative algorithm named *Algorithm BCA* to solve $\hat{\mathcal{P}}$ by combining the convex optimization method (for the optimization of $\mathbf{W}$) and the bisection search method (for the optimization of γ). We first propose a low complexity algorithm to solve $\hat{\mathcal{P}}$ with fixed γ based on the Lagrange dual method. Then we summarize the overall Algorithm BCA and establish its global convergence.

1) Lagrange dual method for solving Problem $\hat{\mathcal{P}}$ with fixed γ: For fixed γ, problem $\hat{\mathcal{P}}$ is convex and thus can be solved by the standard convex optimization method [21]. However, the computation complexity of standard convex optimization method is still quite high as the number of antennas M becomes large. Based on the Lagrange dual method, we exploit the specific structure of the problem to propose a low complexity algorithm which can significantly reduce the computation complexity over the standard convex optimization method.

The Lagrange function of Problem $\hat{\mathcal{P}}$ with fixed γ is

$$\begin{aligned}L\left(\boldsymbol{\mu},\boldsymbol{\nu},\mathbf{W}\right) \ = \ &\gamma+\sum_{n=1}^{N}\nu_n\left(\mathrm{Tr}\left(\overline{\mathbf{\Theta}}_n^{\circ}\mathbf{W}_n\right)-\eta_n\right)+\\ &\sum_{k=1}^{K}\mu_k\Bigg[a_k\mathrm{Tr}\left(\overline{\mathbf{\Theta}}_{\overline{n}_k}^{\circ}\mathbf{W}_{\overline{n}_k}\right)-\\ &\gamma\sum_{n\in\mathcal{B}_k}\mathrm{Tr}\left(\overline{\mathbf{\Theta}}_{k,n}\mathbf{W}_n\right)-c_{k,1}-c_{k,2}\gamma\Bigg],\end{aligned}$$

where $\boldsymbol{\mu}=[\mu_k]_{k=1,...,K}\in\mathbb{R}_+^K$ is the Lagrange multiplier vector associated with the constraint in (12), $\boldsymbol{\nu}=[\nu_n]_{n=1,...,N}\in\mathbb{R}_+^N$ is the Lagrange multiplier vector associated with the constraint in (13), $c_{k,1}=a_k\left(|\mathcal{U}_{\overline{n}_k}|-1\right)$ and $c_{k,2}=\frac{w_k}{\lambda_{\overline{n}_k}P_k}$. Note that $L\left(\boldsymbol{\mu},\boldsymbol{\nu},\mathbf{W}\right)$ is a partial Lagrangian since (14) is kept as a separate constraint. The dual function of Problem $\hat{\mathcal{P}}$ with fixed γ is

$$J\left(\boldsymbol{\mu},\boldsymbol{\nu}\right)\triangleq\max_{\mathbf{W}}L\left(\boldsymbol{\mu},\boldsymbol{\nu},\mathbf{W}\right),\ \text{s.t. (14) is satisfied.} \qquad (15)$$

The corresponding dual problem is

$$\min_{\boldsymbol{\mu},\boldsymbol{\nu}}J\left(\boldsymbol{\mu},\boldsymbol{\nu}\right),\ \text{s.t. }\boldsymbol{\mu}\geq\mathbf{0},\boldsymbol{\nu}\geq\mathbf{0}. \qquad (16)$$

Note that $L\left(\boldsymbol{\mu},\boldsymbol{\nu},\mathbf{W}\right)=\sum_{n=1}^{N}\mathrm{Tr}\left(\mathbf{A}_n\mathbf{W}_n\right)+c^{'}$, where $\mathbf{A}_n=\sum_{k\in\mathcal{U}_n}\mu_k a_k\overline{\mathbf{\Theta}}_{\overline{n}_k}^{\circ}-\gamma\sum_{k\in\overline{\mathcal{U}}_n}\mu_k\overline{\mathbf{\Theta}}_{k,n}+\nu_n\overline{\mathbf{\Theta}}_n^{\circ}$, $\overline{\mathcal{U}}_n=$

$\left\{k:\ \overline{n}_k \neq n, \left(k, \bar{l}_n\right) \in \mathcal{E}\right\}$, and $c^{'}$ is independent of $\mathbf{W}$. Then the maximization problem in (15) can be decomposed into N independent problems as

$$\max_{\mathbf{W}_n} \mathrm{Tr}\left(\mathbf{A}_n \mathbf{W}_n\right),\ \text{s.t. } \mathbf{W}_n - \mathbf{W}_n^2 \succeq \mathbf{0}, \mathbf{W}_n \succeq 0, \qquad (17)$$

for $n = 1, ..., N$. Let $\mathbf{A}_n = \mathbf{U}_n \mathbf{D}_n \mathbf{U}_n^{\dagger}$ be the EVD of $\mathbf{A}_n$ and let $\mathbf{U}_n^*$ denote the eigenvectors corresponding to the positive eigenvalues of $\mathbf{A}_n$. Then for fixed $\boldsymbol{\mu}, \boldsymbol{\nu}$, it can be shown that Problem (17) has a closed form solution given by

$$\mathbf{W}_n^*\left(\boldsymbol{\mu}, \boldsymbol{\nu}\right) = \mathbf{U}_n^* \mathbf{U}_n^{*\dagger}. \qquad (18)$$

Since Problem $\hat{\mathcal{P}}$ with fixed γ is convex, the optimal solution is given by $\mathbf{W}^*\left(\boldsymbol{\mu}^*, \boldsymbol{\nu}^*\right) = \left\{\mathbf{W}_n^*\left(\boldsymbol{\mu}^*, \boldsymbol{\nu}^*\right) : \forall n\right\}$, where $\left(\boldsymbol{\mu}^*, \boldsymbol{\nu}^*\right)$ is the optimal solution of the equivalent dual problem in (16).

The dual function $J\left(\boldsymbol{\mu}, \boldsymbol{\nu}\right)$ is convex and $\nabla J\left(\boldsymbol{\mu}, \boldsymbol{\nu}\right) = \left[\partial_{\mu_1} J\left(\boldsymbol{\mu}, \boldsymbol{\nu}\right), ..., \partial_{\mu_K} J\left(\boldsymbol{\mu}, \boldsymbol{\nu}\right), \partial_{\nu_1} J\left(\boldsymbol{\mu}, \boldsymbol{\nu}\right), ..., \partial_{\nu_N} J\left(\boldsymbol{\mu}, \boldsymbol{\nu}\right)\right]^T$ is a subgradient of $J\left(\boldsymbol{\mu}, \boldsymbol{\nu}\right)$ at $\left(\boldsymbol{\mu}, \boldsymbol{\nu}\right)$, where

$$\begin{aligned}
\partial_{\mu_k} J\left(\boldsymbol{\mu}, \boldsymbol{\nu}\right) &= a_k \mathrm{Tr}\left(\overline{\boldsymbol{\Theta}}_{\overline{n}_k}^{\circ} \mathbf{W}_{\overline{n}_k}^*\left(\boldsymbol{\mu}, \boldsymbol{\nu}\right)\right) - c_{k,2}\gamma \\
&\quad -\gamma \sum_{n \in \mathcal{B}_k} \mathrm{Tr}\left(\overline{\boldsymbol{\Theta}}_{k,n} \mathbf{W}_n^*\left(\boldsymbol{\mu}, \boldsymbol{\nu}\right)\right) - c_{k,1}, \forall k \\
\partial_{\nu_n} J\left(\boldsymbol{\mu}, \boldsymbol{\nu}\right) &= \mathrm{Tr}\left(\overline{\boldsymbol{\Theta}}_n^{\circ} \mathbf{W}_n^*\left(\boldsymbol{\mu}, \boldsymbol{\nu}\right)\right) - \eta_n, \forall n
\end{aligned} \qquad (19)$$

Hence, the standard subgradient based methods such as the subgradient algorithm in [22] or the Ellipsoid method in [21] can be used to solve the optimal solution $\left(\boldsymbol{\mu}^*, \boldsymbol{\nu}^*\right)$ of the dual problem in (16).

Remark 7. For Problem $\hat{\mathcal{P}}$ with fixed γ, the Lagrange dual method is a better choice than other standard convex optimization methods such as interior point method [21] for two reasons: 1) the number of primal variables in this problem is much larger than the number of dual variables ($NM^2 + 1$ versus $K + N$ with $M \gg K$ and $M \gg N$); 2) for fixed Lagrange multipliers $\boldsymbol{\mu}, \boldsymbol{\nu}$, the optimal primal variables $\mathbf{W}_n^*\left(\boldsymbol{\mu}, \boldsymbol{\nu}\right)$ have closed form solution. Hence, although the convergence speed of the simple subgradient and Ellipsoid algorithms for the optimization of dual variables is not as fast as the more complicated algorithms, the overall computation time (to reach a given accuracy) of the Lagrange dual method is much smaller than other standard convex optimization methods.

2) Summary of the Overall Algorithm for solving Problem $\hat{\mathcal{P}}$: Algorithm BCA is summarized in Table II and the global convergence is established in the following theorem.

Theorem 8 (Global Convergence of Algorithm BCA)**.** *For fixed tolerable error ε, Algorithm BCA terminates after $I^{\varepsilon} = \left\lceil log_2\left(\frac{|\overline{\gamma}^{\circ} - \gamma^{\circ}|}{\varepsilon}\right)\right\rceil$ iterations. Let $\mathbf{W}^{\varepsilon} = \mathbf{W}^{(I^{\varepsilon})}, \gamma^{\varepsilon} = \gamma^{(I^{\varepsilon})}$ denote the solution found by Algorithm BCA with tolerable error ε. Then we have $|\gamma^{\varepsilon} - \gamma^{\star}| \leq \varepsilon$. Moreover, as $\varepsilon \to 0$, we have $\mathbf{W}^{\varepsilon} \to \mathbf{W}^{\star}$ and $\gamma^{\varepsilon} \to \gamma^{\star}$.*

Sketch of Proof: Algorithm BCA ensures that $\gamma^{\star} \in [\gamma_{\min}, \gamma_{\max}]$ after each iteration. Moreover, after the i-th iteration, we have $|\gamma_{\max} - \gamma_{\min}| \leq \frac{|\overline{\gamma}^{\circ} - \gamma^{\circ}|}{2^i}$ and $\gamma^{(i)} \in [\gamma_{\min}, \gamma_{\max}]$. Hence, we have $\left|\gamma^{(i)} - \gamma^{\star}\right| \leq \frac{|\overline{\gamma}^{\circ} - \gamma^{\circ}|}{2^i}$, from which Theorem 8 follows immediately. ■

Table II: Algorithm BCA (for solving Problem $\hat{\mathcal{P}}$)

Initialization: Choose $\gamma_{\min} = \gamma^{\circ}$ and $\gamma_{\max} = \overline{\gamma}^{\circ} > \gamma^{\circ}$ such that Problem $\hat{\mathcal{P}}$ with fixed $\gamma = \overline{\gamma}^{\circ}$ is infeasible. Let $i = 1$.
Step 1: Let $\gamma^{(i)} = \frac{\gamma_{\min} + \gamma_{\max}}{2}$. Solve Problem $\hat{\mathcal{P}}$ with fixed $\gamma = \gamma^{(i)}$ using the Lagrange dual method in Section IV-D1.
Step 2: If a feasible solution, denoted by $\mathbf{W}^{(i)}$, is found in Step 1, let $\gamma_{\min} = \gamma^{(i)}$; otherwise, let $\gamma_{\max} = \gamma^{(i)}$.
Step 3: If $|\gamma_{\max} - \gamma_{\min}| \leq \varepsilon$, where $\varepsilon > 0$ is the tolerable error, terminate the algorithm. Otherwise, let $i = i + 1$ and return to Step 1.

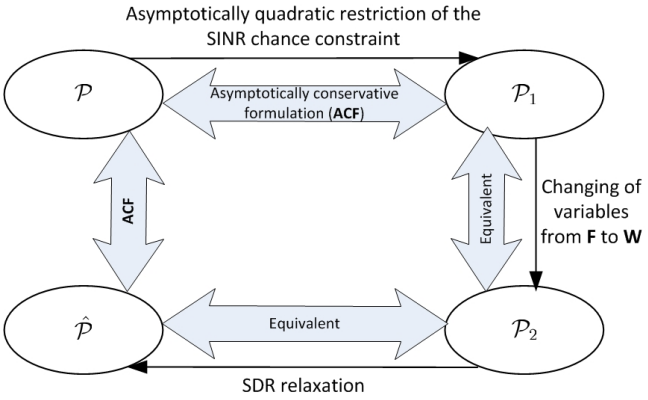


Figure 2: Illustration of the relationship between different problems.

E. Summary of the Overall Solution

The relationship between problems $\mathcal{P}$, $\mathcal{P}_1$, $\mathcal{P}_2$ and $\hat{\mathcal{P}}$ is summarized in Fig. 2 and is elaborated below.

- **Relationship between $\mathcal{P}$ and $\mathcal{P}_1$:** Problem $\mathcal{P}_1$ is obtained by replacing the original SINR chance constraint in (3) with its asymptotically quadratic restriction in (7). Hence, Problem $\mathcal{P}_1$ is an asymptotically conservative formulation of the Problem $\mathcal{P}$, i.e., the solution of $\mathcal{P}_1$ is a feasible solution to the original problem $\mathcal{P}$ for sufficiently large M.
- **Relationship between $\mathcal{P}$ and $\hat{\mathcal{P}}$:** By applying the SDR technique, we obtain Problem $\hat{\mathcal{P}}$ as a relaxation of $\mathcal{P}_1$. By Theorem 6, such relaxation is tight and thus $\hat{\mathcal{P}}$ is equivalent to $\mathcal{P}_1$. Hence, $\hat{\mathcal{P}}$ is also an asymptotically conservative formulation of the Problem $\mathcal{P}$.

Then we summarize the overall solution. First, Algorithm BCA is performed to calculate the optimal solution $\mathbf{W}^{\star}, \gamma^{\star}$ (up to some tolerable error ε) for Problem $\hat{\mathcal{P}}$. Then we calculate the corresponding Tx subspace control $\mathbf{F}^{\star} = \left\{\mathbf{F}_1^{\star}, ..., \mathbf{F}_N^{\star}\right\}$ using EVD $\mathbf{W}_n^{\star} = \mathbf{F}_n^{\star} \mathbf{I}_{S_n^{\star}} \mathbf{F}_n^{\star\dagger}, \forall n$, where $S_n^{\star} = \mathrm{Tr}\left(\mathbf{W}_n^{\star}\right)$. Finally, we output $\mathbf{F}^{\star}, \left\{S_n^{\star}\right\}, \gamma^{\star}$ as an approximate solution to the original Problem $\mathcal{P}$. Note that $\mathbf{F}^{\star}, \left\{S_n^{\star}\right\}, \gamma^{\star}$ is usually a conservative solution in the sense that the actual outage probability of each user k under $\mathbf{F}^{\star}, \left\{S_n^{\star}\right\}, \gamma^{\star}$ is much lower than the maximum allowable outage probability ϵ_k.

Remark 9. In this paper, we focus on the optimization of Tx subspace control $\mathbf{F}$ and subspace dimensions $\{S_n\}$ for fixed power control $\{P_k\}$ (problem $\mathcal{P}$), which is the key of the two stage precoding design for massive MIMO systems. The results and algorithm in this paper also provides a basis

for solving the more complicated joint optimization of Tx subspace control and power control $\{P_k\}$. For example, we may design an alternating optimization (AO) algorithm to solve the joint optimization problem. In each iteration of the AO algorithm, Algorithm BCA is used as a component to find the optimal Tx subspace control for fixed power control, and then the optimal power control is obtained for fixed Tx subspace control.

F. Methods to Improve the Performance Over the Conservative Solution

One common drawback of using the worst-case deterministic constraint to approximate chance constraint is that, the obtained solution is usually too conservative [10]. In the context of the probabilistic SINR constrained beamforming problem, the bisection refinement method has been proposed to mitigate the conservatism due to the deterministic restriction of the probabilistic SINR constraint [10]. This bisection refinement method can also be applied in our problem to improve the performance of the conservative solution obtained in Section IV-E. However, the bisection refinement method in [10] requires solving Problem $\hat{\mathcal{P}}$ for multiple times, which increases the computation complexity. To address this problem, we propose a simple non-iterative refinement method to determine the proper value of δ_k. First, we calculate the simple BD-based subspace control $\mathbf{F}^{\mathrm{BD}}$ using baseline 3 in Section VI. Then, we let $\gamma^{\mathrm{BD}} = \check{\Phi}^{\mathrm{BD}}(\epsilon_k)/w_k, \forall k$, where Φ^{BD} is the empirical CDF of the SINR of user k, denoted by $\mathrm{SINR}_k^{\mathrm{BD}}$, under BD-based subspace control $\mathbf{F}^{\mathrm{BD}}$, and $\check{\Phi}^{\mathrm{BD}}$ is the inverse function of Φ^{BD}. Note that $\Pr\left\{\mathrm{SINR}_k^{\mathrm{BD}} \geq w_k\gamma^{\mathrm{BD}}\right\} \approx 1-\epsilon_k$. Finally, δ_k is chosen as the largest number that satisfies (7) with $\mathbf{F} = \mathbf{F}^{\mathrm{BD}}$ and $\gamma = \gamma^{\mathrm{BD}}$. Simulations show that the non-iterative refinement method achieves almost the same performance as the much more complicated bisection refinement method.

V. Implementation Considerations

In this section, we discuss various implementation issues, such as how to obtain the statistical and real-time CSI, the signaling overhead and the impact of CSI error. In the following analysis, we assume that the spatial channel correlation matrices $\mathbf{\Theta}$ change every T time slots and the rank of each spatial correlation matrix is R.

A. Slow Statistical CSI Acquisition

Under typical scenario, the spatial channel correlation matrices change after a few seconds and the time slot duration is in the order of millisecond, i.e., T is usually in the order of a few thousands. For example, in urban areas with a carrier frequency of 2GHz and a user speed of 3km/h, the spatial channel correlation matrices remain unchanged for about 22 seconds according to the measurement results reported in [23]. Hence, the statistical CSI acquisition can be done at a much slower timescale compared to the real-time CSI acquisition. Specifically, we propose a *rank deficient oracle approximating shrinkage (OAS) estimator* which is able to obtain an accurate estimation of the spatial correlation matrix using only $T_p = O(R) \ll M$ independent channel samples.

Rank deficient OAS estimator:

Input: T_p independent channel samples $\mathbf{h}(j) \in \mathbb{C}^M, j = 1, ..., T_p$ of a random vector $\mathbf{h} \in \mathbb{C}^M$ with covariance matrix $\mathbf{\Theta}$.

Step 1: Using Gram–Schmidt process to obtain a semi unitary matrix $\mathbf{U} \in \mathbb{C}^{M\times R}$ whose columns form an orthogonal basis for $\mathbf{h}(j) \in \mathbb{C}^M, j = 1, ..., T_p$.

Step 2: Form T_p independent channel samples with reduced dimension $\mathbf{h}^{'}(j) = \mathbf{U}^{\dagger}\mathbf{h}(j) \in \mathbb{C}^R, j = 1, ..., T_p$. Note that $\mathbf{h}^{'}(j)$'s can be viewed as realizations of a random vector $\mathbf{h}^{'} \in \mathbb{C}^R$ with covariance matrix $\mathbf{U}^{\dagger}\mathbf{\Theta}\mathbf{U}$.

Step 3: Use the OAS estimator in [24] with independent samples $\mathbf{h}^{'}(j), j = 1, ..., T_p$ to obtain an estimation $\hat{\mathbf{\Theta}}^{'}$ of the covariance matrix of $\mathbf{h}^{'}$.

Step 4: Output the estimation of the covariance matrix of $\mathbf{h}$ as: $\hat{\mathbf{\Theta}} = \mathbf{U}\hat{\mathbf{\Theta}}^{'}\mathbf{U}^{\dagger}$.

Then the statistical CSI acquisition is summarized as follows. Within each T time slots, choose T_p time slots such that the gap between any two selected time slots is larger than channel coherent time. Then each BS transmits M dedicated pilot symbols in each of the selected T_p time slots for the users to estimate the spatial correlation matrices using the rank deficient OAS estimator. Finally, user k sends the non-zero eigenvalues and the corresponding eigenvectors of $\mathbf{\Theta}_{k,l}, \forall l$ to the associated BS.

B. Real-time CSI Acquisition

The dimension of the effective channel vector of a user in the n-th user cluster is equal to S_n (the rank of the n-th Tx subspace control variable $\mathbf{F}_n$), which is $O(1)$ and is much less than M. To estimate the effective CSI $\widetilde{\mathbf{H}}_n$ of the n-th user cluster, the pilot symbols need to be precoded using the n-th Tx subspace control variable $\mathbf{F}_n$ since the effective CSI $\widetilde{\mathbf{H}}_n$ is restricted in the subspace spanned by $\mathbf{F}_n$. As a result, we can simultaneously transmit N pilot symbols at a time for N user clusters because the Tx subspace control $\mathbf{F}$ can also mitigate the inter-cluster interference in the pilot transmission stage. The total number of orthogonal pilot symbols required for effective CSI estimation $\widetilde{\mathbf{H}}_n, n = 1, ..., N$ is $\max_n S_n = O(1)$. Hence, by treating $\mathbf{F}_n$ as part of the effective channel, the estimation of the effective CSI $\widetilde{\mathbf{H}}_n$ is similar to the conventional channel estimation problem in a single-cell small-scale MIMO system and the conventional CSI signaling method in LTE can be used to estimate the effective CSI.

C. Statistical and Real-time CSI Signaling Overhead

The CSI signaling overhead includes the pilot symbol overhead (the average number of orthogonal pilot symbols for real-time and statistical CSI estimation) and the uplink CSI feedback overhead (the average number of feedback vectors with different dimensions). For simplicity, assume that each BS is associated with K_0 users, each user is interfered by $L_0 - 1$ neighbor BSs, and $S_n = S, \forall n$. Then the real-time CSI signaling overhead of the proposed scheme is "S PS, K_0 $\mathbb{C}^S$" per time slot per cell, which means that, the

proposed scheme requires transmitting S independent pilot symbols and feed backing K_0 complex channel vectors of dimension S per time slot per cell. For each spatial correlation matrix, the corresponding user only needs to send R non-zero eigenvalues and the corresponding eigenvectors to the BS. Hence, the statistical CSI signaling overhead of the proposed scheme is "$\frac{MT_p}{T}$ PS, $\frac{K_0L_0R}{T}$ $\mathbb{C}^M$, $\frac{K_0L_0}{T}$ $\mathbb{R}^R$" per time slot per cell, which means that, the proposed scheme requires transmitting $\frac{MT_p}{T}$ independent pilot symbols, feed backing $\frac{K_0L_0R}{T}$ complex vectors of dimension M (eigenvectors of spatial correlation matrices), and feed backing $\frac{K_0L_0}{T}$ real vectors of dimension R (eigenvalues of spatial correlation matrices) per time slot per cell. As a comparison, the CSI signaling overhead of the conventional single-cell MU-MIMO precoding scheme is "M PS, K_0 $\mathbb{C}^M$" per time slot per cell. Since $T \gg T_p = O(R) = O(L_0) = O(1)$, the overall CSI signaling overhead of the proposed scheme is significantly smaller as will be shown in the simulations.

D. Backhaul Signaling Overhead

Suppose there are a total number of N_L links (including both direct links and cross links) in the network. The BSs need to send R non-zero eigenvalues and the corresponding eigenvectors of every spatial correlation matrix to the central node within each T time slots. Hence, the backhaul signaling overhead is "$\frac{N_LR}{T}$ $\mathbb{C}^M$, $\frac{N_LR}{T}$ $\mathbb{R}$" per time slot (i.e., $\frac{N_LR}{T}$ complex vectors of dimension M and $\frac{N_LR}{T}$ real numbers per time slot). Since $T \gg R$ is usually true for massive MIMO systems, the backhaul signaling overhead can be greatly reduced compared to the cooperative MIMO or centralized coordinated MIMO where the BSs need to send N_L complex channel vectors of dimension M to the central node at each time slot.

E. Summary of Practical Real-time Implementation Aspects

In the proposed two stage precoding technique, there are four components (steps) working at different timescales.

Step 1 (Slow Statistical CSI Acquisition): First, the BSs need to obtain the spatial channel correlation matrices using the method in Section V-A, and then send them to the central node.

Step 2 (Tx Subspace Control Optimization): With the obtained statistical CSI from the BSs, the central node calculates the optimal Tx subspace control variables $\{\mathbf{F}_n\}$ using Algorithm BCA and sends $\mathbf{F}_n$ to BS $\bar{l}_n$ (which is the BS associated with the n-th user cluster).

Step 3 (Real-time CSI Acquisition): By treating $\mathbf{F}_n$ as part of the effective channel, the estimation of the real-time effective CSI $\widetilde{\mathbf{H}}_n$ is similar to the conventional channel estimation problem in a single-cell small-scale MIMO system.

Step 4 (Inner Precoding): At each time slot, BS $\bar{l}_n$ performs inner precoding using the obtained effective CSI $\widetilde{\mathbf{H}}_n$.

Both the statistical CSI acquisition and the Tx subspace control optimization can be done at a much slower timescale compared to the time slot duration. Hence, the proposed solution is not sensitive to the backhaul signaling latency delay and computational delay caused by the Tx subspace control optimization. On the other hand, the effective CSI acquisition and inner precoding per cluster is similar to that in the single-cell small-scale MIMO system and thus they can be implemented in real-time at each time slot. Note that in practice, it is impossible to obtain perfect effective CSI at each BS due to the channel estimation/feedback error. The impact of imperfect effective CSI on the performance is similar to that in the conventional small-scale MIMO systems with ZF precoding. As a result, we can also employ MMSE techniques for the inner precoder to mitigate the effects of CSI errors. The theoretical analysis of the effect of imperfect effective CSI on the two stage subspace constrained precoding is left as an interesting future work.

Remark 10 (Extension to Frequency Selective Fading Channels). For clarity, we assume flat fading channel in this paper. In the wideband systems (such as OFDM) with frequency selective channels, the spatial correlation matrices are approximately identical on different subcarriers [25], [26]. As a result, the same Tx subspace control $\mathbf{F}$ can be applied to different subcarriers and the results and algorithm can also be extended to the wideband systems. The extension to wideband systems depends on the specific QoS requirements. If we require that the SINR of each user on each subcarrier must be larger than a threshold with high probability, then the proposed solution can be directly applied. However, if we require that the rate of each user (rate summed over all subcarriers) must be larger than a threshold with high probability, then the extension may be non-trivial.

VI. Simulation Results

A. Simulation Setup

Consider a massive MIMO cellular system with 19 cells, where a reference cell is surrounded by two tiers of (interfering) cells. The inter-site distance is 500m. In each cell, there are K_0 users and 2 uniformly distributed hotspots (user clusters). Each hotspot contains K_c users and the other $K_0 - 2K_c$ users are uniformly distributed within the cell. Each BS is equipped with M antennas. The path gains $L_{k,l}$'s are first generated using the path loss model ("Urban Macro NLOS" model) in [27]. Then the spatial correlation matrices $\mathbf{\Theta}_{k,l}$'s are generated according to the local scattering model in [6], [7]. Specifically, the correlation between the i,m-th channel coefficients of $\mathbf{h}_{k,l}$ is given by

$$[\mathbf{\Theta}_{k,l}]_{i,m} = \int_{\varphi_{k,l}^{\min}}^{\varphi_{k,l}^{\max}} \psi_{k,l}(\varphi)\, e^{j\varpi_l^{i,m}(\varphi)} d\varphi, \qquad (20)$$

where φ denotes the AoD, $\psi_{k,l}(\varphi)$ denotes the power gain for the path specified by the AoD φ and $\varpi_l^{i,m}(\varphi)$ is the phase difference between the i-th and m-th BS antennas on the path specified by φ. In the simulations, $\varpi_l^{i,m}(\varphi)$ in (20) is generated using a uniform linear array with the antenna spacing equal to half wavelength, and $\psi_{k,l}(\varphi) = \overline{\psi}_{k,l}, \forall \varphi \in \left[\varphi_{k,l}^{\min}, \varphi_{k,l}^{\max}\right]$, where $\overline{\psi}_{k,l}$ is chosen such that $\mathrm{Tr}(\mathbf{\Theta}_{k,l}) = ML_{k,l}$ and $\varphi_{k,l}^{\min}, \varphi_{k,l}^{\max}$ are randomly generated such that 1) $\left|\varphi_{k,l}^{\max} - \varphi_{k,l}^{\min}\right| = \frac{\pi}{6}$, 2) if user k and user k' belong to the

same hotspot, we have $\varphi_{k,l}^{\min} = \varphi_{k',l}^{\min}, \varphi_{k,l}^{\max} = \varphi_{k',l}^{\max}$. We assume the spatial channel correlation matrices $\mathbf{\Theta}$ change every 1000 time slots.

In the *basic simulation setup*, the per BS transmit power is $P_b = 10$dB and the other system parameters are set as $K_0 = 7$, $K_c = 3$, $M = 40$, $\epsilon_k = 0.05, \forall k$ and $w_k = L_{k,b_k}, \forall k$; the transmit powers of the users are given by $P_k = \frac{P_b}{K_0}, \forall k$. In the simulations, we will change the system parameters in the basic setup to study the impact of different system parameters on the performance. We only evaluate the performance of the reference cell to avoid "unsymmetric edge effects".

We compare the performance of the proposed solution with the following baselines.

- **Baseline 1 (FFR)**: Fractional frequency reuse (FFR) [28] is applied to suppress the inter-cell interference. There are 6 subbands, 3 of which are used for the inner zone and 3 are used for the outer zone. The radius for the inner zone is set as 150m. In each cell, ZF beamforming is used to serve the users on each subband.
- **Baseline 2 (Clustered CoMP)**: 3 neighbor BSs form a BS cluster and employ cooperative ZF [5] to simultaneously serve all the users within the BS cluster.
- **Baseline 3 (JSDM-PGP)**: This is the two-stage precoding scheme proposed in [8], where the per-group processing (PGP) with approximate BD is employed at each BS. In the approximate BD, the pre-beamforming matrices $\mathbf{F}$ are chosen to satisfy $\mathbf{F}_n^{\dagger}\mathbf{U}_{k,\bar{l}_n}^{\star} = 0, \forall n \in \mathcal{B}_k, \forall k$, where $\mathbf{U}_{k,l}^{\star} \in \mathbb{U}^{M \times r_{k,l}^{\star}}$ is a semi-unitary matrix collecting the dominant eigenvectors of $\mathbf{\Theta}_{k,l}$. The number $r_{k,l}^{\star}$ of dominate eigenvectors is chosen such that $\lambda_{k,l}^{dB}\left(r_{k,l}^{\star}+1\right) \leq -20\text{dB} < \lambda_{k,l}^{dB}\left(r_{k,l}^{\star}\right)$, where $\lambda_{k,l}^{dB}(m)$ is the m-th largest eigenvalue of $\mathbf{\Theta}_{k,l}$ in dB. Simulations show that such choice of $r_{k,l}^{\star}$ can achieve a good performance. Both ZF inner precoder and regularized zero-forcing (RZF) [2] inner precoder will be simulated.

B. Verify the SINR Satisfaction Probability

First, we verify that the original QoS requirement in (3) is satisfied by the proposed solution under finite M. Note that the QoS requirement in (3) means that the *minimum SINR satisfaction probability* $\min_k \Pr\{\text{SINR}_k \geq \gamma\}$ (conditioned on a given spatial correlation matrices $\mathbf{\Theta}$) is larger than some target value $1-\epsilon_k$. Fig. 3 shows the histograms of the minimum SINR satisfaction probabilities over different realizations of $\mathbf{\Theta}$ under the basic simulation setup ($M = 40$). To obtain the histograms, we generated 100 realizations of $\mathbf{\Theta}$. Then, for each realization of $\mathbf{\Theta}$, the minimum SINR satisfaction probability is numerically evaluated using 10000 randomly generated realizations of instantaneous CSI $\mathbf{H}$. Fig. 3 validates that the proposed solution indeed achieve a minimum SINR satisfaction probability no less than the required target 95%. The "probability gap" in Fig. 3 refers to the difference between the mean minimum SINR satisfaction probability and the target SINR satisfaction probability. Without using the refinement methods, the proposed solution is conservative, e.g., the "probability gap" is 0.042. However, with the simple non-iterative refinement method, the proposed solution become "tighter" (i.e., the "probability gap" is smaller), and with the more complicated bisection refinement method in [10], the proposed solution can be even tighter.

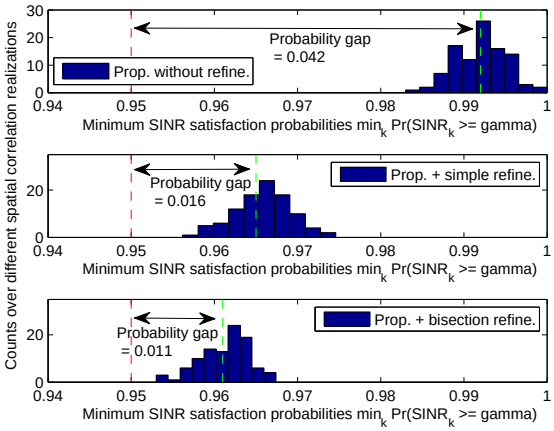


Figure 3: Histograms of the minimum SINR satisfaction probabilities under the basic simulation setup. The red dash line indicates the target SINR satisfaction probability and the green dash line indicates the mean minimum SINR satisfaction probability.

C. Comparison of Outage Throughput Performance under Different System Parameters

In Fig. 4, we plot the outage throughput of different schemes under the basic simulation setup, where the outage throughput of user k is defined as the maximum data rate that can be achieved with a probability no less than $1-\epsilon_k$. For baseline 2, the 3 cooperative BSs need to exchange CSI and payload data, and thus there is CSI delay when the backhaul latency is not zero. When there is CSI delay, the outdated CSI is related to the actual CSI by the autoregressive model in [29] with the following parameters: the user speed is 3 km/h; the carrier frequency is 2GHz. For baseline 3, ZF inner precoder is simulated. It can be seen that the performance of the proposed scheme with the simple non-iterative refinement method is close to that with the more complicated bisection refinement method in [10]. The outage throughput of the proposed scheme is larger than baseline 1 and baseline 3. Although the performance of baseline 2 is promising at zero backhaul latency, the performance quickly degrades at 10ms backhaul latency.

In Fig. 5, we change the number of antennas to $M = 64$. The other simulation setup is the same as the basic simulation setup. Compared to Fig. 4 with $M = 40$ antennas, the performance of all schemes improves as the number of antennas increases. However, the performance improvement of the proposed scheme and baseline 3 is larger. This is not surprising since the proposed scheme and baseline 3 are specifically designed for massive MIMO systems with spatially correlated channel.

In Fig. 6, we change the number of users in each cell to $K_0 = 8$ and the number of users per cluster to $K_c = 4$.

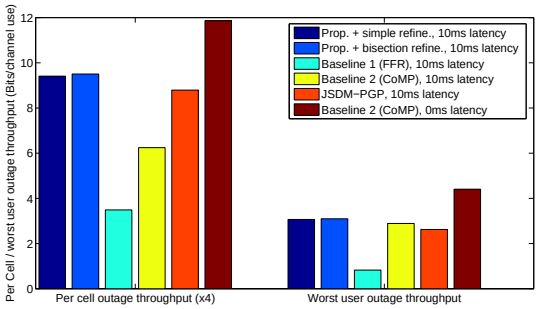


Figure 4: Outage throughput comparisons of different schemes under the basic simulation setup.

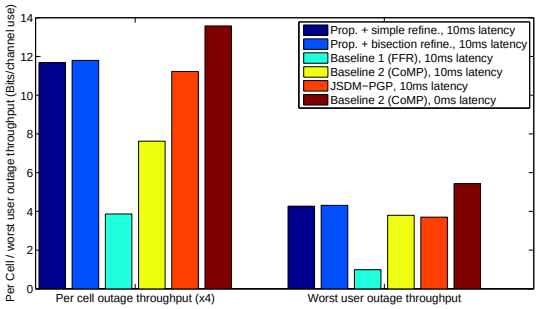


Figure 5: Outage throughput comparisons of different schemes. The number of antennas is $M = 64$ and the other simulation setup is the same as the basic simulation setup.

The other simulation setup is the same as the basic simulation setup. In this case, the users in each cell concentrate in the two hotspots. Compared to Fig. 4, the performance of both the proposed scheme and baseline 3 are improved. This shows that the proposed scheme and baseline 3 favor the scenario when the users in each cell concentrate in a few user clusters.

In Fig. 7, we consider heterogeneous QoS requirements, where ϵ_k is chosen uniformly between $[0.05, 0.15]$. The other simulation setup is the same as the basic simulation setup. Compared to Fig. 4 with homogeneous QoS requirement, the performance gain of the proposed scheme over baseline 1 and

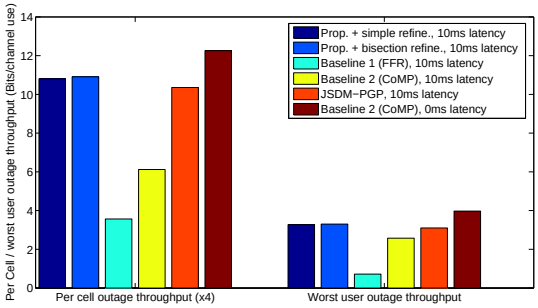


Figure 6: Outage throughput comparisons of different schemes. The number of users in each cell is $K_0 = 8$ and the number of users per cluster is $K_c = 4$. The other simulation setup is the same as the basic simulation setup.

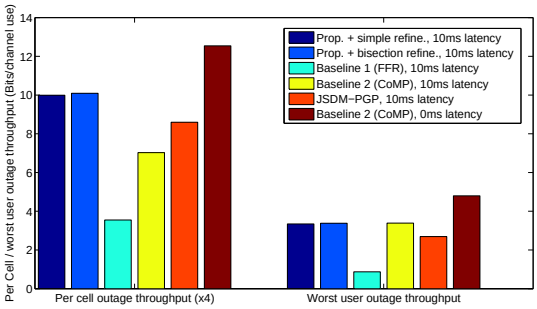


Figure 7: Outage throughput comparisons of different schemes with heterogeneous QoS requirements. The other simulation setup is the same as the basic simulation setup.

3 becomes larger. This shows that the proposed solution can better handle the heterogeneous QoS requirements.

Under all simulation setups, the proposed scheme outperforms baseline 1 and baseline 3, as well as baseline 2 under the practical scenario with 10ms backhaul latency. These results demonstrate the superior performance and the robustness of the proposed optimization based subspace constrained precoding w.r.t. signaling latency in backhaul.

D. Impact of Inner Percoders and CSI Errors

In this paper, we consider ZF inner precoding because 1) the ZF inner precoding is more tractable; and 2) it has been shown in [1] that ZF precoding is close to optimal for massive MIMO systems with perfect CSIT. To study the impact of different inner precoders, we compare the performance of the ZF inner precoding with the RZF inner precoding under different *outer precoders* $\mathbf{F}$, where the outer precoder refers to either the proposed optimization based Tx subspace control matrix or the pre-beamforming matrix in baseline 3. In RZF inner precoder, the regularization factor is fixed to $\frac{K_0}{S_0 P}$ as in [8], where S_0 is the total dimension of the outer precoders associated with the user clusters in the reference cell. To study the impact of CSI errors, the CSIT at each BS is modeled as follows:

$$\hat{\mathbf{h}}_{k,l} = \widetilde{\mathbf{h}}_{k,l} - \mathbf{e}_{k,l},$$

where the CSI error $\mathbf{e}_{k,l}$ is assumed to be a complex Gaussian vector with zero mean and covariance matrix $\sigma_e^2\mathbf{I}$, and σ_e^2 denotes the error variance. In Fig. 8, we plot the outage throughput of different schemes versus the error variance σ_e^2 under the basic simulation setup. It can be seen that the performance of ZF inner precoding under the same outer precoder is similar to that of RZF inner precoding. The performance of all schemes degrades at a similar rate as the CSI error increases. In all cases, the proposed optimization based outer precoder outperforms the BD-based out percoder in baseline 3.

E. Implementation Cost

Table III compares the implementation cost for different schemes in terms of the computational complexity (CPU time), the (real-time and statistical) CSI signaling overhead, and

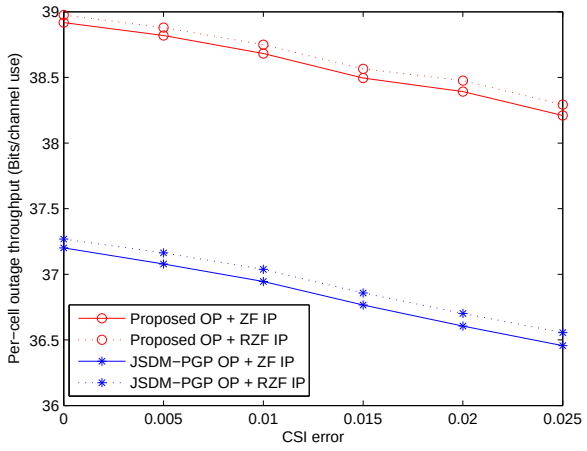


Figure 8: An illustration of the impact of inner precoding and CSI errors under the basic simulation setup. The abbreviation "OP" stands for outer precoding and "IP" stands for inner precoding.

	CPU time	Backhaul signaling	Real-time CSI signaling	Statistical CSI signaling
Proposed	0.13 s	0.13 $\mathbb{C}^{40}$	9 PS, 7 $\mathbb{C}^{9}$	0.8 PS, 0.12 $\mathbb{C}^{40}$
FFR	0.06 s	0	40 PS, 7 $\mathbb{C}^{40}$	0
Ideal CoMP	1.29 s	16 $\mathbb{C}^{40}$	40 PS, 7 $\mathbb{C}^{120}$	0
JSDM-PGP	0.04 s	0.13 $\mathbb{C}^{40}$	9 PS, 7 $\mathbb{C}^{9}$	0.8 PS, 0.12 $\mathbb{C}^{40}$

Table III: Comparison of the per time slot MATLAB computational time and per time slot per cell signaling overhead of different schemes under the basic simulation setup. For the proposed scheme and Baseline 1, the statistical CSI $\mathbf{\Theta}$ is obtained using the method described in Section V-A with $T_p = 20$. The meaning of the CSI signaling overhead notation in the table is also explained in Section V-A.

the backhaul signaling overhead caused by the exchange of (real-time or statistical) CSI between the central node and BSs. It can be seen that the overall implementation cost of the proposed scheme and baseline 3 is the lowest among all schemes.

VII. Conclusion

We propose a two-stage subspace constrained precoding scheme for massive MIMO cellular systems. The MIMO precoder is partitioned into *inner precoder* (for intra-cluster spatial multiplexing gain) and *Tx subspace control* matrix (for inter-cluster interference control). We formulate the Tx subspace control as a Quality of Service (QoS) optimization problem and propose a novel bi-convex approximation approach to produce efficiently computable bi-convex approximation of this QoS optimization problem. Analysis shows that the proposed solution is robust to backhaul signaling latency and can significantly reduce the CSI signaling overhead. Moreover, simulations verify that the proposed design achieves significant performance gain compared with various state-of-the-art baselines under various backhaul signaling latency.

Appendix

A. Proof of Theorem 3

Note that $I_k = \sum_{n\in\mathcal{B}_k} M\overline{\mathbf{z}}_{k,\bar{l}_n}^{\dagger}\mathbf{\Theta}_{k,\bar{l}_n}^{1/2}\mathbf{F}_n\mathbf{G}_n\mathbf{P}_n\mathbf{G}_n^{\dagger}\mathbf{F}_n^{\dagger}\mathbf{\Theta}_{k,\bar{l}_n}^{1/2}\overline{\mathbf{z}}_{k,\bar{l}_n}$, where $\overline{\mathbf{z}}_{k,\bar{l}_n} \triangleq \frac{1}{\sqrt{M}}\mathbf{z}_{k,\bar{l}_n} \in \mathbb{C}^M$ has i.i.d. complex entries of zero mean and variance $\frac{1}{M}$. It can be verified that the inner ZF precoder $\mathbf{G}_n$ satisfies $\mathbf{G}_n\mathbf{G}_n^{\dagger} = \widetilde{\mathbf{H}}_n^{\dagger}\left(\widetilde{\mathbf{H}}_n\widetilde{\mathbf{H}}_n^{\dagger}\right)^{-1}\mathrm{diag}\left(\left(\widetilde{\mathbf{H}}_n\widetilde{\mathbf{H}}_n^{\dagger}\right)^{-1}\right)^{-1}\left(\widetilde{\mathbf{H}}_n\widetilde{\mathbf{H}}_n^{\dagger}\right)^{-1}\widetilde{\mathbf{H}}_n$, from which it follows that

$$\left\|\mathbf{G}_n\mathbf{P}_n\mathbf{G}_n^{\dagger}\right\| \le M\overline{P}_n\widetilde{\lambda}_{n,\max}\left\|\widetilde{\mathbf{H}}_n^{\dagger}\left(\widetilde{\mathbf{H}}_n\widetilde{\mathbf{H}}_n^{\dagger}\right)^{-2}\widetilde{\mathbf{H}}_n\right\| \le \frac{\overline{P}_n\widetilde{\lambda}_{n,\max}}{\widetilde{\lambda}_{n,\min}},$$

where $\widetilde{\lambda}_{n,\max}$ and $\widetilde{\lambda}_{n,\min}$ are respectively the largest and the smallest eigenvalue of $\frac{1}{M}\widetilde{\mathbf{H}}_n\widetilde{\mathbf{H}}_n^{\dagger}$. It follows from Assumption 2 that $\limsup_{M\to\infty}\left\|\frac{1}{M}\mathbf{H}_n\mathbf{H}_n^{\dagger}\right\| < \infty$ with probability 1 [15], where $\mathbf{H}_n = \left[\mathbf{h}_{k,\bar{l}_n}\right]_{k\in\mathcal{U}_n}^{\dagger} \in \mathbb{C}^{|\mathcal{U}_n|\times M}$. Since $\widetilde{\lambda}_{n,\max} = \left\|\frac{1}{M}\widetilde{\mathbf{H}}_n\widetilde{\mathbf{H}}_n^{\dagger}\right\| \le \left\|\frac{1}{M}\mathbf{H}_n\mathbf{H}_n^{\dagger}\right\|$, $\widetilde{\lambda}_{n,\max}$ is uniformly bounded with probability 1. On the other hand, there exists $\widetilde{\lambda}_0 > 0$ such that, we have $\widetilde{\lambda}_{n,\min} \ge \widetilde{\lambda}_0$ uniformly on M; otherwise, it will contradict with the assumption that $\mathbf{F}$ is feasible. It follows from the above analysis that $\left\|\mathbf{G}_n\mathbf{P}_n\mathbf{G}_n^{\dagger}\right\|$ is uniformly bounded on M with probability 1. Together with Assumption 2, $\left\|\mathbf{\Theta}_{k,\bar{l}_n}^{1/2}\mathbf{F}_n\mathbf{G}_n\mathbf{P}_n\mathbf{G}_n^{\dagger}\mathbf{F}_n^{\dagger}\mathbf{\Theta}_{k,\bar{l}_n}^{1/2}\right\|$ is also uniformly bounded on M with probability 1. Then by [15, Lemma 4], we have $I_k - \overline{I}_k \overset{a.s}{\to} 0$, where

$$\begin{aligned}\overline{I}_k &= \sum_{n\in\mathcal{B}_k}\mathrm{Tr}\left(\mathbf{\Theta}_{k,\bar{l}_n}^{1/2}\mathbf{F}_n\mathbf{G}_n\mathbf{P}_n\mathbf{G}_n^{\dagger}\mathbf{F}_n^{\dagger}\mathbf{\Theta}_{k,\bar{l}_n}^{1/2}\right)\\ &\le \sum_{n\in\mathcal{B}_k}\mathrm{Tr}\left(\mathbf{G}_n\mathbf{P}_n\mathbf{G}_n^{\dagger}\right)\mathrm{Tr}\left(\mathbf{F}_n^{\dagger}\mathbf{\Theta}_{k,\bar{l}_n}\mathbf{F}_n\right) = \hat{I}_k,\end{aligned}$$

from which Theorem 3 follows immediately.

B. Proof of Theorem 4

Let $\widetilde{\mathbf{H}}_{-k} = \left[\widetilde{\mathbf{h}}_{k',\overline{n}_k}\right]_{k'\in\mathcal{U}_{\overline{n}_k}\setminus\{k\}}^{\dagger} \in \mathbb{C}^{(|\mathcal{U}_{\overline{n}_k}|-1)\times S_{\overline{n}_k}}$ and consider the singular value decomposition (SVD): $\widetilde{\mathbf{H}}_{-k} = \widetilde{\mathbf{U}}\widetilde{\mathbf{D}}\widetilde{\mathbf{V}}^{\dagger}$, where $\widetilde{\mathbf{U}} \in \mathbb{U}^{(|\mathcal{U}_{\overline{n}_k}|-1)\times(|\mathcal{U}_{\overline{n}_k}|-1)}$, $\widetilde{\mathbf{D}} \in \mathbb{R}^{(|\mathcal{U}_{\overline{n}_k}|-1)\times(|\mathcal{U}_{\overline{n}_k}|-1)}$, $\widetilde{\mathbf{V}} \in \mathbb{U}^{S_{\overline{n}_k}\times(|\mathcal{U}_{\overline{n}_k}|-1)}$. Let $\widetilde{\mathbf{V}}_c \in \mathbb{U}^{S_{\overline{n}_k}\times(S_{\overline{n}_k}-|\mathcal{U}_{\overline{n}_k}|+1)}$ be the orthogonal complement of $\widetilde{\mathbf{V}}$, i.e., $\widetilde{\mathbf{V}}_c^{\dagger}\widetilde{\mathbf{V}} = \mathbf{0}$. Using similar proof as that of [30], it can be shown that $\left|\mathbf{h}_{k,b_k}^{\dagger}\mathbf{F}_{\overline{n}_k}\mathbf{g}_k\right|^2 = \mathbf{h}_{k,b_k}^{\dagger}\mathbf{F}_{\overline{n}_k}\widetilde{\mathbf{V}}_c\widetilde{\mathbf{V}}_c^{\dagger}\mathbf{F}_{\overline{n}_k}^{\dagger}\mathbf{h}_{k,b_k}$. Hence, conditioned on $\widetilde{\mathbf{H}}_{-k}$, $\widetilde{\mathbf{V}}_c$ is fixed, and thus $\mathbf{e} \triangleq \widetilde{\mathbf{V}}_c^{\dagger}\mathbf{F}_{\overline{n}_k}^{\dagger}\mathbf{h}_{k,b_k} \sim \mathcal{CN}(\mathbf{0},\mathbf{Q})$ is a complex Gaussian vector with covariance $\mathbf{Q} = \widetilde{\mathbf{V}}_c^{\dagger}\mathbf{F}_{\overline{n}_k}^{\dagger}\mathbf{\Theta}_{\overline{n}_k}^{\circ}\mathbf{F}_{\overline{n}_k}\widetilde{\mathbf{V}}_c$. Then conditioned on $\widetilde{\mathbf{H}}_{-k}$, $\left|\mathbf{h}_{k,b_k}^{\dagger}\mathbf{F}_{\overline{n}_k}\mathbf{g}_k\right|^2 = \overline{\mathbf{e}}^{\dagger}\mathbf{Q}\overline{\mathbf{e}}$, where $\overline{\mathbf{e}}$ is a standard complex Gaussian vector. Applying the Bernstein-type inequality based method in [9], [10], we have

$$\Pr\left\{\left|\mathbf{h}_{k,b_k}^{\dagger}\mathbf{F}_{\overline{n}_k}\mathbf{g}_k\right|^2 \ge \frac{w_k\gamma}{P_k}\left(\hat{I}_k+1\right)\middle|\widetilde{\mathbf{H}}_{-k}\right\} \ge 1-\epsilon_k, \quad (21)$$

if the following sufficient condition is satisfied

$$\mathrm{Tr}\left(\overline{\mathbf{Q}}\right) - \sqrt{2\delta_k}\sqrt{\left\|\overline{\mathbf{Q}}\right\|_F^2} \ge \frac{w_k\gamma}{\lambda_{\overline{n}_k}P_k}\left(\hat{I}_k+1\right), \quad (22)$$

where $\overline{\mathbf{Q}} = \frac{1}{\lambda_{\overline{n}_k}}\mathbf{Q}$. It is easy to see that all the eigenvalues of $\overline{\mathbf{Q}}$ must be no more than 1. As a result, we have $\left\|\overline{\mathbf{Q}}\right\|_F^2 \leq \mathrm{Tr}\left(\overline{\mathbf{Q}}\right)$ and thus (22) can be satisfied if

$$\mathrm{Tr}\left(\overline{\mathbf{Q}}\right) - \sqrt{2\delta_k}\sqrt{\mathrm{Tr}\left(\overline{\mathbf{Q}}\right)} \geq \frac{w_k\gamma}{\lambda_{\overline{n}_k}P_k}\left(\hat{I}_k + 1\right). \quad (23)$$

It is easy to verify that (23) can be satisfied if

$$\begin{aligned}\left(1 - \frac{\sqrt{2\delta_k}}{\sigma_k}\right)\mathrm{Tr}\left(\overline{\mathbf{Q}}\right) &\geq \frac{w_k\gamma}{\lambda_{\overline{n}_k}P_k}\left(\hat{I}_k + 1\right), \\ \mathrm{Tr}\left(\overline{\mathbf{Q}}\right) &\geq \sigma_k^2. \end{aligned} \quad (24)$$

It can be shown that $\mathrm{Tr}\left(\overline{\mathbf{Q}}\right) \geq \sum_{j=|\mathcal{U}_{\overline{n}_k}|}^{S_{\overline{n}_k}} \overline{\lambda}_j$, where $\overline{\lambda}_j$ is the j-th largest eigenvalue of $\mathbf{F}_{\overline{n}_k}^{\dagger}\overline{\mathbf{\Theta}}_{\overline{n}_k}^{\circ}\mathbf{F}_{b_k}$. Using the fact that $\overline{\lambda}_j \leq 1, \forall j$, we have $\sum_{j=|\mathcal{U}_{\overline{n}_k}|}^{S_{\overline{n}_k}} \overline{\lambda}_j \geq \mathrm{Tr}\left(\overline{\mathbf{\Theta}}_{\overline{n}_k}^{\circ}\mathbf{F}_{\overline{n}_k}\mathbf{F}_{\overline{n}_k}^{\dagger}\right) - |\mathcal{U}_{\overline{n}_k}| + 1$. Hence $\mathrm{Tr}\left(\overline{\mathbf{Q}}\right) \geq \mathrm{Tr}\left(\overline{\mathbf{\Theta}}_{\overline{n}_k}^{\circ}\mathbf{F}_{\overline{n}_k}\mathbf{F}_{\overline{n}_k}^{\dagger}\right) - |\mathcal{U}_{b_k}| + 1$. Note that (7) is obtained from (24) by replacing $\mathrm{Tr}\left(\overline{\mathbf{Q}}\right)$ with $\mathrm{Tr}\left(\overline{\mathbf{\Theta}}_{b_k}^{\circ}\mathbf{F}_{b_k}\mathbf{F}_{b_k}^{\dagger}\right) - |\mathcal{U}_{b_k}| + 1$. Hence, (7) ensures that (24) is satisfied. Clearly, the above relationship between (21-24) and (7) holds for any $\widetilde{\mathbf{H}}_{-k}$. As a result, (7) also ensures that the unconditional version of (21) in (6) is satisfied.

C. Proof of Theorem 6

It is easy to see that $\mathbf{W}^{\star}$ must also be the optimal solution of $\hat{\mathcal{P}}$ with fixed $\gamma^{\star}$, which is a convex problem w.r.t. $\mathbf{W}^{\star}$. According to the analysis in Section IV-D1, there exist Lagrange multipliers $\boldsymbol{\mu}^{\star}$ and $\boldsymbol{\nu}^{\star}$ such that $\mathbf{W}^{\star}$ maximizes the corresponding Lagrange function $L\left(\boldsymbol{\mu}^{\star}, \boldsymbol{\nu}^{\star}, \mathbf{W}\right)$ in (15). The optimal solution $\mathbf{W}^{\star}$ of the maximization problem in (15) with $(\boldsymbol{\mu}, \boldsymbol{\nu}) = (\boldsymbol{\mu}^{\star}, \boldsymbol{\nu}^{\star})$ is given by (18) with $(\boldsymbol{\mu}, \boldsymbol{\nu})$ in $\mathbf{A}_n$ equal to $(\boldsymbol{\mu}^{\star}, \boldsymbol{\nu}^{\star})$. From (18), it is easy to see that $\mathbf{W}_n^{\star} - \mathbf{W}_n^{\star 2} = \mathbf{0}, \forall n$, from which Theorem 6 follows immediately.

REFERENCES

[1] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[2] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195 – 202, Jan. 2005.

[3] A. Gershman, N. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 62–75, 2010.

[4] G. Foschini, K. Karakayali, and R. Valenzuela, "Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency," *IEE Proceedings on Communications*, vol. 153, no. 4, pp. 548 – 555, Aug. 2006.

[5] O. Somekh, O. Simeone, Y. Bar-Ness, A. Haimovich, and S. Shamai, "Cooperative multicell zero-forcing beamforming in cellular downlink channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3206–3219, 2009.

[6] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels," *IEEE J. Select. Areas Commun.*, vol. 20, no. 3, pp. 550–560, 2002.

[7] M. Zhang, P. Smith, and M. Shafi, "An extended one-ring MIMO channel model," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2759–2764, 2007.

[8] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing - the large-scale array regime," *IEEE Trans. Info. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct 2013.

[9] I. Bechar, "A Bernstein-type inequality for stochastic processes of quadratic forms of gaussian variables," 2009, preprint. [Online]. Available: http://arxiv.org/abs/0909.3595

[10] K.-Y. Wang, T.-H. Chang, W.-K. Ma, A.-C. So, and C.-Y. Chi, "Probabilistic SINR constrained robust transmit beamforming: A Bernstein-type inequality based conservative approach," in *in Proc. IEEE ICASSP 2011*, May 2011, pp. 3080–3083.

[11] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Select. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug 2002.

[12] D. Chizhik, J. Ling, P. Wolniansky, R. Valenzuela, N. Costa, and K. Huber, "Multiple-input-multiple-output measurements and modeling in manhattan," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 321–331, Apr 2003.

[13] *Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced*, Geneva, Switzerland, Rep. ITU-R M.2135- 1, Dec. 2009.

[14] W. Qianya and H. Yang, "A switched diversity scheme for massive mimo systems," *International Journal of Antennas and Propagation*, vol. 2014, 2014.

[15] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Info. Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012.

[16] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Select. Areas Commun.*, vol. 31, no. 2, pp. 160–171, 2013.

[17] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Select. Areas Commun.*, vol. 31, no. 2, pp. 264–273, 2013.

[18] A. Adhikary and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming and user grouping," 2013. [Online]. Available: http://arxiv.org/abs/1305.7252

[19] A. Ben-Tal and A. Nemirovski, "On safe tractable approximations of chance-constrained linear matrix inequalities," *Math. Oper. Res.*, vol. 1, pp. 1–25, Feb. 2009.

[20] A. M. Tulino and S. Verdu, *Random Matrix Theory and Wireless Communications*. Foundations and Trends in Communications and Information Theory, vol. 1, no. 1, June 2004.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[22] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," 2003. [Online]. Available: http://www.stanford.edu/class/ee392o

[23] I. Viering, H. Hofstetter, and W. Utschick, "Spatial long-term variation in urban, rural and indoor environments," in *Proceedings of the 5th COST*, vol. 273, 2002.

[24] Y. Chen, A. Wiesel, Y. Eldar, and A. Hero, "Shrinkage algorithms for MMSE covariance estimation," *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5016–5029, Oct 2010.

[25] S. Chandran, *Adaptive Antenna Arrays: Trends and Applications*. Springer, 2004.

[26] A. Sadek, W. Su, and K. Liu, "Transmit beamforming for space-frequency coded MIMO-OFDM systems with spatial correlation feedback," *IEEE Trans. Commun.*, vol. 56, no. 10, pp. 1647–1655, Oct. 2008.

[27] *Technical Specification Group Radio Access Network; Further Advancements for E-UTRA Physical Layer Aspects*, 3GPP TR 36.814. [Online]. Available: http://www.3gpp.org

[28] H. Lei, L. Zhang, X. Zhang, and D. Yang, "A novel multi-cell OFDMA system structure using fractional frequency reuse," *in Proc. IEEE Int. Symp. Personal, Indoor Mobile Radio Commun.*, pp. 1–5, Sep. 2007.

[29] K. Baddour and N. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1650–1662, 2005.

[30] P. Li, D. Paul, R. Narasimhan, and J. Cioffi, "On the distribution of SINR for the MMSE MIMO receiver and performance analysis," *IEEE Trans. Info. Theory*, vol. 52, no. 1, pp. 271–286, 2006.